Determine the relationship between House Prices and local Income Level

Hongzhuo Yin Spring 2021

Abstract

Real estate market is one of the most important fractions of the financial industry, and housing price has always been the spotlight of discussion for centuries. In this paper, we want to discuss the relationship between housing prices and the local personal income per capita. The other explanatory variables including price level, GDP growth rate (compound over 2 years), and population net flow of a specific area. The discussion will be cross-sectional comparing data of different Metropolitan Statistical areas. Since the Housing Price Index from Federal Housing Finance Agency was calibrated in 2000, so we chose a recent year 2018 where all data were collected and still considered closely related to current situation. Through this study, a weak positive relationship between personal income per capita was displayed with linear regression model.

I. Introduction

Ever since the financial crisis in 2008, the house prices seem to rise un-stoppable. While there was trade war, business cycles, pandemic, none of them seems to actually stop the house prices climbing in this past 10 years. What actually determine the house prices? As one of the largest purchases in people's life, house price is related to every single person. To a nation, the real estate market is one of the most important property markets. Housing itself provide great investment value. Banks give out mortgages and sell mortgage-backed securities into financial market. As a large industry, real estate could connect to construction, furnishing, utility, household appliances, etc., in addition to its large property tax contribution to the local government. Because of these, house is the most special "daily good" we would consider. Thus, we want to determine what could affect the house price level would see this problem from market level's supply-demand relation as well as the general economic trend.

From this research paper, I wish to determine the relationship between house prices and local income level. With our basic understanding of supply and demand relationship of economics, it would as well explain the real estate market. However, as one of the most important factors of economics that's closely related to the general publics, many researchers and scholars spent time and resources discovering the logistics behind the housing price. Many indexes were invented to express housing price more directly: Housing Price Index shows the change in housing price, House-price-to-Income-Ratio express the ratio between household's house price and income, house affordability index measures how affordable the houses are based on mortgage offer and income together, etc. All these indexes are useful helping people directly understand the trend and visualize the data. However, these indexes are not enough to find direct relationship between house price and other independent variables.

From the history of learning real estate market, we could see several factors that plays important role in determining house prices. Local income level is one of the most important. It represents the purchase power of the market, and it is a main factor affecting the demand. Therefore, the house prices as the result of supply and demand curve, would be dependent under the effect of local income level. Through this paper, we will set house prices as the dependent variable, and the local income level as the primary independent variable. With the simple regression model, we will measure both house prices and the individual income level by Metropolitan statistical area and compare cross-sectionally. And in the multiple-regression model, we will bring in other control variables that also determines the house price.

II. Literature Review (about 2 to 2.5 pages long)

Dennis R.Capozza, Patric H. Hendershott, Charlotte Mack, and Christopher J. Mayer from National Bureau of Economic Research have published a paper in 2002 on the determinants of real house price dynamics. They discussed the serial correlation and mean reversion parameters and how they are varied across different metro-areas. They determined that the variation in house prices across different areas are more than just local economies, but also other factors such as development speed, population growth rate, growth in real income, construction cost (which is highly dependent on whether the area is developing or well-developed). It is interesting that they have discussed the information cost and the mobility of the housing market, the possibility of new investors entering the market and cause housing price changes. While this paper was mainly focusing on 1980s housing prices, it has pointed out the difference among cities' development stage and their different responses with such reason. Thus, as we are analyzing across different metropolitan statistical areas, we ensure the randomness.

Dr. Joshua Gallin from Federal Reserve Board has published a paper "the Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets" in 2006, where he discussed the long run cointegration between house price and income. He argues that by the data, these two factors are not co-integrated. With small sample size, the traditional times-series test does not work well as cross panel method. The author collected data for 95 metro areas over 27 years, and showed that with such data and stronger test, it still does not show clear evidence of cointegration between house price and the income level. Dr. Gallin has used population, stock market, construction wage, PCE deflator, as the explanatory variables along with the primary variable of personal income per capita. He has analyzed the supply and the demand of the housing market. The demand of the local housing market would depend on buyer's income (personal income per capita), amount(population), wealth, user cost of housing; and the supply would depend on the price of the housing, cost of production(construction), where the user cost are costs related to mortgage, taxes, fees, net capital gain and the cost of maintenance. From here, Dr. Gallin is breaking down the supply-demand into quantifiable variables. From this equation, we could see that the house price is related to the income per capita, however Dr. Gallin was showing there might not necessary exist a cointegration. While in this paper I'm not determining the correction between times series process in long term, his method is useful in term of how what variables I should pay attention to and factors that I could use as explanatory variables.

Martijn I. Droes and Alex van de Minne has published a paper in 2017 discussing the housing price from a unique perspective. They analyzed whether the determinants of housing price would change over time, and they used Netherland' 200 years of transaction data. While from many papers we could found the cost of construction determines the supply, and consumer's amount and income determine the demand. However, how much do each of them actually impact the housing price are unknown. The authors of the paper "Do the Determinants of House Prices Change over Time? Evidence from 200 Years of Transactions Data" has determined what has to the value of housing. Before 1900s, construction cost was the main fraction, whereas after industrial evolution, income, mortgage and interest rates played an important role. Then, with baby booming and the effect of World War II, population and reconstruction play became the main factor. Nowadays, location and the attached value worth more than the house itself. In the long run, with the improvement of technology and the seasonal effect, housing prices determinants are changing continuously. In my data, I would not only consider the current determinants: income level, GDP, price level, but also population migration that would impact the housing price through affecting the total demand.

In my paper, I wish to find the cross-sectional relationship between housing price and personal income level, as well as other factors like GDP growth, price level, population flow, and how these could be expressed as a quantitative relationship to the housing price. In such a way, we could update on the current determinants of the housing price, as well as determine which variable would have the largest effect on the overall housing price. Therefore, I wish this paper could help the reader have a better view of what is the current key determinants of the housing price.

III. Data

In additional to the primary independent variable: income level, we have found other independent variables that are affecting our dependent variable house prices. Despite the buyer's purchasing power—which is income level, the quantity of consumer is also important. Nowadays, population are migrating between cities and areas. Human resource inflow would bring wealth into an area. For example, California, Silicon Valley's house price rocketed over the past 30 years, where large technology companies located their headquarters and attracted over thousands of high-paying families migrated into the area. Thus, we would also include the **net migration** of each metropolitan statistical area as a control variable. In addition to the buyers, local economy plays an important role in real estate market as well. The **price level** of different area is different. If we ignore the role of price level, we might have overestimated the effect of the primary independent variables, since income level and price level are closely correlated. Since real estate are considered an investment in financial industry, thus, we would like to consider local **GDP growth rate** over 2 years as a control variable. This variable would be a general intuition about investors' expectation and reflected on the house prices.

Since the object of this paper is on the house price in United States determined by metropolitan statistical area, I have drawn most of my economic data from the official website of Bureau of Economic Analysis under United States Department of Commerce. From this source, I have drawn data on **per capita personal income** by MSA, measured in dollar, and used natural logarithm on it to make it reflect on change in percentage; **Regional Price Parities** by MSA, for 2018, is the price level indicator organized by the bureau of economic analysis, reflecting the difference across state and metropolitan areas. It is measured as index compare to the whole united states as 100; **GDP growth rate** by MSA, which is compounded over 2018, measured in percentage. I have obtained the house price index from Federal Housing Finance Agency, which organize **House Price Index (HPI)** and offer dataset for historical data. I extracted the Purchase-only index from 2018. Since our purpose of this study is to analyze the house price in the perspective of affordability for purchase, so the purchase-only index would better serve. It was estimated using only the Sales Price, eliminated the effect of other transactions and the effect appraisal price. Also, since HPI is measured quarterly, so we could use the Q4 data as the end-of-year data to compare to other annually measured variables. To eliminate the seasonal components, we used the seasonal adjusted index. In addition, I have obtained the **net Flow** which is measured in number of individuals migrated from area A to B. I used area A as the base and determined the total net population flow to each of the area, so the unit is in a thousand of individuals. I obtained the data from United States Census Bureau, which were collected from 2014-2018 American Community Survey. Then, I wanted to determine the factor of total **population** in effect of the local housing prices, the data was also from Census Bureau. At last, I obtained an index measuring local **economy freedom**, which is a score range from 1 to 10 measuring the local tax policies, government spending, labor market freedom, and such qualitative measure of the local economy condition. This data is obtained from Reason Foundation, a non-profit think tank in California.

Through the data analysis, I have used the Core based Statistical areas (CBSA) code to match the dataset for each metropolitan statistical area, we have picked the 100 largest MSA as our sample. The list of MSA could be found in the appendix. Below is our scatterplot for 2018 natural logarithm of personal income vs. HPI for different MSA, which shows weak correlation between the two variables. The outliers with HPI over 5 is Denver, Colorado where had great housing price increment in the past 20 years (Since the HPI used year 2000 as the base year). The three outliers with high personal income are

San Francisco, San Jose, and Stamford, where large technology companies located their headquarters and employed high income people, driving up the average personal income.

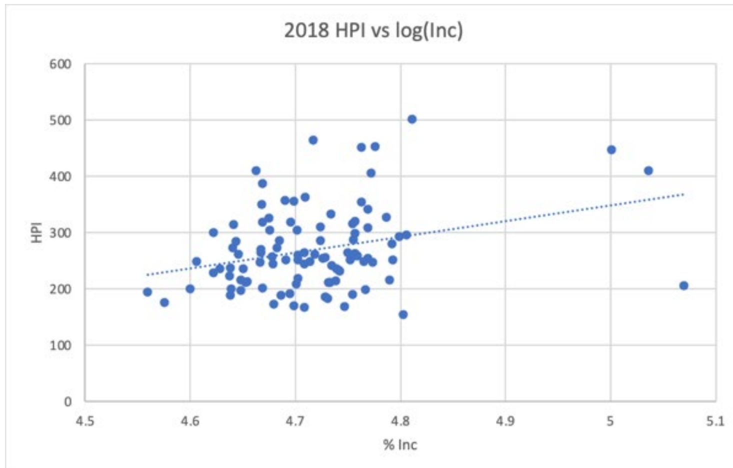

Figure 1- Scatterplot of Natural Logarithm of Income per capita vs. HPI

The following table is a summary of each variable.

Table 1. Varaible Description

Variable Name	Description	Year	Units	Source
HPI	House Price Index	2018	1 point	Federal House Finance Agency
log (Inc)	Natural Logarithm of Personal Income per capita	2018	percentage	Bureau of Economic Analysis
PL	Price Level	2018	1 point	Bureau of Economic Analysis
GDPgrowth	GDP growth rate compounded over 2 years	2017-2019	percentage	Bureau of Economic Analysis
netFlow	Population net flow for a specific metropolitan statistical area in	2014-2018	Number of people	United States Census Bureau

	thousands of people			
Pop	Total population estimated	2018	Partition of total population of US in MSA, in percentage	United States Census Bureau
EconFrdm	Economic Freedom score	2012	Score from 0 to 10	Reason Foundation

Table 2. Varaible Descriptive Statistics

Variable Name	Observation	Mean	Standard Deviation	Minimum	Maximum
HPI	100	269.73	72.41	155.19	502.28
log (Inc)	100	4.72	0.078	4.56	5.07
PL	100	96.96	8.81	87.5	133.9
GDPgrowth	100	4.68	1.20	1.4	7.5
netFlow	100	0.2656	21.652	-128.89	48.585
Pop	100	0.6764	0.9680	0.04007	6.8531
EconFrdm	100	6.689	0.7231	5.109	8.548

The above two tables described all the variables and their descriptive statistics. From table 2 we could see that the netFlow has very high standard deviation, thus high variance as well. For Personal Income per capita, price level, GDP growth rate, and net flow, there are more than 100 metropolitan statistical area has been observed. However, since we only choose sample size to be 100 for the housing price index, we have matched the corresponding areas and made the observation size to be 100 for each of the variable. Since we are finding the relationship with HPI, so it’s better to trim the other variable sample size rather than leaving some entries of the dependent variable empty.

Before Constructing the regression model, we would need to test the five Gauss Markov Assumptions and Normality for the Classical Linear Model. According to the book, we have the following assumptions: (The following assumptions are directly quoted from the book).

1. **Linear in Parameters:**

The model can be written as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$,

Where $\beta_0, \beta_1, \dots, \beta_k$ are known are the unknown parameters (constants) of interest and u is an unobserved random error or disturbance term.

2. **Random Sampling:**

We are using the data of metropolitan areas available by the Federal House Finance Agency on Housing Price Index; thus, we have not manipulated which area to use, thus this sampling is random.

3. **No Perfect Collinearity:**

None of the control variable is a linear function for another, as we could have check it by the software as well.

4. **Zero Conditional Mean:**

We could not claim that there are no other factors in u such that it could influence the house price. Thus, we should observe this with caution.

5. **Homoskedasticity:**

We could not claim that the error u has the same variance given any value of the explanatory variable, as u may contain factors that could influence the value of explanatory variable.

6. **Normality**

The population error u is independent of the explanatory variable and is normally distributed with zero mean and variance. We could not conclude this based on our data availability.

III. Results (2 to 3 pages)

<u>Simple Regression (1)</u>

For the simple regression, we assume the function has format:

$$y = \beta_0 + \beta_1 x + u$$

Where in our regression it would looks like:

$$HPI = \beta_0 + \beta_1 \log(Inc) + u$$

Estimated by the software, we could obtain the following equation:

$$HPI = -1055.41 + 280.89 \log(Inc)$$

Having R-squared as 0.0921, denoting week correlation between housing price index and the natural logarithm of personal income per capita. Since β_1 is positive, it shows a positive linear relationship between personal income and the housing price of an area. Because we have used natural logarithm on personal income per capita, we could interpret the function as: for every 1% increase in the personal income per capita, there would be a 280.89 increase on the Housing Price Index for that metropolitan statistical Area. For this model, as we are testing the significance level, we noticed that log(Inc) is significant at 1% level, the intercept is significant at 5% level.

We could see that the simple regression model could not explain the dependent variable well, so we would need other independent control variables to form a multiple regression model, and better approach the result.

Multiple Regression Model (2)

For a Multiple Regression model, the function for this Housing Price index would have the following format:

$$HPI = \beta_0 + \beta_1\, log(Inc) + \beta_2 PL + \beta_3 GDPgrowth + \beta_4 netFlow + \beta_5 Pop + \beta_6 EconFrdm + u$$

Using the software we could get an estimation function as the following:

$$HPI = -549.40 + 65.34 \log(Inc) + 2.517PL + 24.78GDPgrowth + 0.10001netFlow - 1.197Pop + 22.66EconFrdm$$

Which have R-square as 0.3994. It’s a better estimation compare to the previous simple regression model. We could see that all the coefficients are having positive sign, meaning that there’s positive correlation between the control variables and the dependent variable. 1% increase in personal income per capita in the area would result in 65.34 increase in Housing Price Index, while holding other factors constant. 1 point increase in the price level would result in 2.517 increase in Housing Price index, holding other variables constant. 1 percent increase in GDP growth rate compound over 2 years would result in 24.78 increase in Housing Price index while holding other variables constant. We could see that the coefficient for net Flow is small, which means it has relative smaller effect on the total housing prices. If 1000 more people migrate to the area would lead the Housing Price Index increase by 0.10001 while holding other variables constant. It's surprising that the total population is having a negative correlation with HPI, where increase of 1% of MSA respect to total US population would result in decrease of 1.197 for HPI. At last, we could see the positive correlation, having 1 point increase in Economic Freedom Score would result in 22.66 points increase in HPI. In this case, price level is significant at 5% level, while both GDP growth rate and economic freedom score are significant at 1% level.

Multiple Regression Model (3)

In this model we will take out the net Flow, Population, and Economic Freedom Score, thus forms the following function:

$$HPI = \beta_0 + \beta_1\, log(Inc) + \beta_2 PL + \beta_2 GDPgrowth + +\beta_5 EconFrdm + u$$

Then we use the software and plug in data, we could obtain the following estimation function:

$$HPI = -658.04 + 132.79\log(Inc) + 1.767PL + 27.74GDPgrowth$$

It has R-squared as 0.3498, which is less than the model 2. It represents that having the three variables is helping to form better estimation. We could also see that the intercept became smaller, and the coefficient in front of log(Inc) and PL became smaller, which means they are having less impact on the Housing Price index, while GDP growth rate is having larger impact on the housing price index. Since increase in net Flow means increase in total population, so total wealth, thus having greater GDP growth rate. Thus, it is expected to see GDP growth is overestimated by covering net Flow's impact on Housing Price Index. In this model, price level is significant at 10% level, GDP growth rate is significant at 1% level.

Multiple Regression Model (4)

In this model we will take out the net Flow and Population only, adding the Economic Freedom Score back. Since GDP growth rate is related to the population inflow and increase in population and net Flow would increase GDP, thus GDP growth rate as well. Thus, by taking out the net Flow and Population, we would expect an overestimation on GDP growth rate.

The function would have format:

$$HPI = \beta_0 + \beta_1\, log(Inc) + \beta_2 PL + \beta_2 GDPgrowth + \beta_3 netFlow + u$$

And as we put it into the software and run estimation, we could obtain the following estimated function:

$$HPI = -516.44 + 58.80 \log(Inc) + 2.436GDPgrowth + 25.20GDPgrowth + 23.11netFlow$$

The R-square is 0.3983, which is less than Model 2. We could see that compare to Model 2, the coefficient on log(Inc), price level decreased, and on GDP growth rate, Economic Freedom Score increased. This again verifies that population is positively correlated to GDP data. In this model, price level is significant at 5% level, while GDP growth rate and economic freedom score are significant at 1% level. Comparing to model 3, R-squared number is greater, showing that the variable of economic freedom score is significant in terms of explaining HPI.

Multiple Regression Model (5)

In this model we will take out both GDP growth rate and net flow, so we would form the following model:

$$HPI = \beta_0 + \beta_1\, log(Inc) + \beta_2 PL + \beta_5 Pop + \beta_6 EconFrdm + u$$

Using software, we could obtain the following estimated function:

$$HPI = -170.94 + 106.61\, log(Inc) + 3.421PL + 0.2418Pop + 31.03EconFrdm$$

This model has R-squared value of 0.2340, decreased significantly from our second model. It means for 1 percent change of personal income, HPI would increase 106.61, for 1 point increase in price level, HPI would increase 3.421. HPI would also 0.2418 for 1% increase in population and 31.03 for 1 point increase in economic freedom score. Among these variables, by the t-test, price level is significant at 5% level and economic freedom score is significant at 1% level.

Overall, we have the results of the model in the following table:

Table 3. Regression Model Summary

Dependent Variable: Housing Price Index					
Independent Variables	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
log(Inc)	280.89***	65.34	132.79	58.80	106.61
PL		2.5169**	1.767*	2.436**	3.421**
GDPgrowth		24.777***	27.74***	25.20***	
netFlow		0.1001			
Pop		-1.196			0.2418
EconFrdm		22.659***		23.11***	31.03***
Intercept	-1055.41** (420.36)	-549.40 (450.88)	-658.04 (451.47)	-516.44 (439.61)	-170.94 (492.27)
R-Squared	0.0921	0.3994	0.3498	0.3983	0.2340

*Significant at 10%, **5%, ***1%

IV. Extensions (2 to 3 pages)

In this section, we will discuss the significance of each variable and models using F-test.

Using the t-test, we were able to find individual variables that’s insignificant. To see their joint significance, we would use F-test to find their jointly significance. Model (2) consists of all the variables, so regardless of what variables we are testing, we could use Model 2 as the unrestricted model. First, comparing Model (2) and Model (3), we are testing the significance of net Flow, population, and economic freedom score, we have having the following hypothesis:

$$H_0: \beta_4 = 0, \beta_5 = 0, \beta_6 = 0$$

$$H_1: H_0\ is\ false$$

Model (3) here is the restricted model. Using the residual sum of squares of Model (2) and Model (3), having q=3 exclusion restriction variables, and there were total of k+1=7 parameters. Having 100 observant, the degree of freedom for unrestricted model is 100-7=93. Here we follow the equation:

$$F = \frac{({R_{ur}}^2 - {R_r}^2)/q}{(1 - {R_{ur}}^2)/(n-k-1)} = \frac{(0.3994 - 0.3498)/3}{(1 - 0.3994)/(100 - 6 - 1)} = 2.5601$$

From the F score table, we could find that $F_{3,93}$= 4.01 at 1% significance, $F_{3,93} = 2.71$ at 5% significance, $F_{3,93} = 2.15$ at 10% significance. Since 2.15 < 2.5601 < 2.71, we failed to reject the null hypothesis at 5% level. We could conclude that net Flow, population, and economic freedom score are jointly significant at 10% level.

Comparing Model (2) and Model (4), we could obtain the jointly significance of net Flow and population, which we would have hypothesis:

$$H_0:\ \beta_4 = 0, \beta_5 = 0$$

$$H_1: H_0\ is\ false$$

Here Model (4) is the restricted model, having q=2 as the exclusion restriction variables. The other number regarding the unrestricted model is same as the previous F-test, so we have

$$F = \frac{({R_{ur}}^2 - {R_r}^2)/q}{(1 - {R_{ur}}^2)/(n-k-1)} = \frac{(0.3994 - 0.3983)/2}{(1 - 0.3994)/(100 - 6 - 1)} = 0.0852$$

Now from the table we could obtain that $F_{2,93}$= 4.85 at 1% significance, $F_{2,93} = 3.10$ at 5% significance, $F_{2,93} = 2.36$ at 10% significance. Since we have 0.0852< 2.36, so net Flow and population are jointly insignificant in our model.

After using the two F-tests as well as considering the individual significance level of independent variables from the t-test, we could notice that net Flow, Population and personal income are not significant in our measuring of the Housing Price Index. Thus, we formed the Model (6) as the following.

Alternative Model: Model (6)

Since we have noticed that log(Inc) was not significant at 10% level in previous 3 multi-variable linear regression, we are taking it out here and only consider price level, GDP growth rate, and economic freedom score. The function would have format:

$$HPI = \beta_0 + \beta_2 PL + \beta_2 GDPgrowth + \beta_5 EconFrdm + u$$

Using the software we could get an estimation function as the following:

$$HPI = 281.96 + 2.820PL + 24.81GDPgrowth + 24.24EconFrdm$$

This model has R-squared as 0.3964, which is a little bit less than model (2) and model (4), showing that log(Inc) is not as significant a factor as we expected it would be. By taking out the factor of personal income, we observed a significant increase in coefficient of price level, which could be seen as an overestimation of the price level. In this model, all the variables, as well as the intercept, are significant at 1% level.

V. Conclusions (About 2 to 3 paragraphs; 1 page)

Throughout our analysis of the relationship between HPI and other independent variables, we have noticed that local personal income is only significant in our simple regression model. Thus, while our hypothesis was not fully supported by the multi-variable regression model. Since we could not show that we meet the normality assumption, the result might be biased as well.

From our alternative Model (6), we could see it is having R-squared value close to our unrestricted model 2, which showed that price level, GDP growth rate, and economic freedom score are having jointly positive correlation with Housing Price Index. GDP growth rate represents the area's economic development trend in the past 2 years, and the economic freedom score would show the effect of local policies on the housing prices, these factors were commons seen as the driven factor for housing prices as well. However, as we also noticed, that the local personal income level only weakly correlated to the housing prices, which is also shown in the scattered plot. In addition, I used personal income level as the indicator, which does not well represent the income level of a household. For single-family houses, if data of household income could be available by MSA, it would generate better result.

Overall, from the result of this paper, further research could be conducted focusing on further secondary variables. Also, since our data is cross sectional comparing different variables of the same year, further research could be done using longitudinal data focusing on specific area, monitor the change of housing prices along with the change of other variables such as price level, GDP per capita, and population growth. Due to housing prices’ close relationship with the financial market and economic condition, housing prices under special conditions (such as economic regression) would also be interesting for further research.

Appendix (not counted in page limits, give list of countries, cities, provide STATA output)

List of MSAs:

10420	Akron, OH
10580	Albany-Schenectady-Troy, NY
10740	Albuquerque, NM
10900	Allentown-Bethlehem-Easton, PA-NJ
11244	Anaheim-Santa Ana-Irvine, CA (MSAD)
12060	Atlanta-Sandy Springs-Alpharetta, GA
12420	Austin-Round Rock-Georgetown, TX
12540	Bakersfield, CA
12580	Baltimore-Columbia-Towson, MD
12940	Baton Rouge, LA
13820	Birmingham-Hoover, AL
14260	Boise City, ID
14454	Boston, MA (MSAD)
14860	Bridgeport-Stamford-Norwalk, CT
15380	Buffalo-Cheektowaga, NY
15764	Cambridge-Newton-Framingham, MA (MSAD)
15804	Camden, NJ (MSAD)
15980	Cape Coral-Fort Myers, FL
16700	Charleston-North Charleston, SC
16740	Charlotte-Concord-Gastonia, NC-SC
16984	Chicago-Naperville-Evanston, IL (MSAD)
17140	Cincinnati, OH-KY-IN
17460	Cleveland-Elyria, OH
17820	Colorado Springs, CO
17900	Columbia, SC
18140	Columbus, OH
19124	Dallas-Plano-Irving, TX (MSAD)
19430	Dayton-Kettering, OH
19740	Denver-Aurora-Lakewood, CO
19804	Detroit-Dearborn-Livonia, MI (MSAD)
20994	Elgin, IL (MSAD)
21340	El Paso, TX
31140	Louisville/Jefferson County, KY-IN
32820	Memphis, TN-MS-AR
33124	Miami-Miami Beach-Kendall, FL (MSAD)
33340	Milwaukee-Waukesha, WI
33460	Minneapolis-St. Paul-Bloomington, MN-WI
33874	Montgomery County-Bucks County-Chester County, PA (MSAD)
34980	Nashville-Davidson--Murfreesboro--Franklin, TN
35004	Nassau County-Suffolk County, NY (MSAD)
35084	Newark, NJ-PA (MSAD)
35300	New Haven-Milford, CT
35380	New Orleans-Metairie, LA
35614	New York-Jersey City-White Plains, NY-NJ (MSAD)
35840	North Port-Sarasota-Bradenton, FL
36084	Oakland-Berkeley-Livermore, CA (MSAD)
36420	Oklahoma City, OK
36540	Omaha-Council Bluffs, NE-IA
36740	Orlando-Kissimmee-Sanford, FL
37100	Oxnard-Thousand Oaks-Ventura, CA
37964	Philadelphia, PA (MSAD)
38060	Phoenix-Mesa-Chandler, AZ
38300	Pittsburgh, PA
38900	Portland-Vancouver-Hillsboro, OR-WA
39300	Providence-Warwick, RI-MA
39580	Raleigh-Cary, NC
40060	Richmond, VA
40140	Riverside-San Bernardino-Ontario, CA
40380	Rochester, NY
40900	Sacramento-Roseville-Folsom, CA
41180	St. Louis, MO-IL
41620	Salt Lake City, UT
41700	San Antonio-New Braunfels, TX
41740	San Diego-Chula Vista-Carlsbad, CA

22744	Fort Lauderdale-Pompano Beach-Sunrise, FL (MSAD)	41884	San Francisco-San Mateo-Redwood City, CA (MSAD)
23104	Fort Worth-Arlington-Grapevine, TX (MSAD)	41940	San Jose-Sunnyvale-Santa Clara, CA
23224	Frederick-Gaithersburg-Rockville, MD (MSAD)	42644	Seattle-Bellevue-Kent, WA (MSAD)
23420	Fresno, CA	44700	Stockton, CA
23844	Gary, IN (MSAD)	45060	Syracuse, NY
24340	Grand Rapids-Kentwood, MI	45104	Tacoma-Lakewood, WA (MSAD)
24660	Greensboro-High Point, NC	45300	Tampa-St. Petersburg-Clearwater, FL
24860	Greenville-Anderson, SC	46060	Tucson, AZ
25540	Hartford-East Hartford-Middletown, CT	46140	Tulsa, OK
26420	Houston-The Woodlands-Sugar Land, TX	46520	Urban Honolulu, HI
26900	Indianapolis-Carmel-Anderson, IN	47260	Virginia Beach-Norfolk-Newport News, VA-NC
27260	Jacksonville, FL	47664	Warren-Troy-Farmington Hills, MI (MSAD)
28140	Kansas City, MO-KS	47894	Washington-Arlington-Alexandria, DC-VA-MD-WV (MSAD)
28940	Knoxville, TN	48424	West Palm Beach-Boca Raton-Boynton Beach, FL (MSAD)
29404	Lake County-Kenosha County, IL-WI (MSAD)	48620	Wichita, KS
29820	Las Vegas-Henderson-Paradise, NV	48864	Wilmington, DE-MD-NJ (MSAD)
30780	Little Rock-North Little Rock-Conway, AR	49180	Winston-Salem, NC
31084	Los Angeles-Long Beach-Glendale, CA (MSAD)	49340	Worcester, MA-CT

Regressions:

Model 1

Regression Statistics	
Multiple R	0.30346519
R Square	0.09209112
Adjusted R Sq	0.08282675
Standard Erro	69.3504233
Observations	100

ANOVA

	df	SS	MS	F	Significance F
Regression	1	47807.9158	47807.9158	9.94034777	0.00214619
Residual	98	471329.159	4809.48121		
Total	99	519137.074			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1055.4125	420.35834	-2.5107448	0.013684	-1889.6	-221.22508	-1889.6	-221.22508
X Variable 1	280.895944	89.0932249	3.15283171	0.00214619	104.093339	457.698549	104.093339	457.698549

Model 2:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.63201578
R Square	0.39944395
Adjusted R Sq	0.3606984
Standard Erro	57.899701
Observations	100

ANOVA

	df	SS	MS	F	Significance F
Regression	6	207366.165	34561.0274	10.3094145	9.8745E-09
Residual	93	311770.91	3352.37537		
Total	99	519137.074			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-549.403546	450.878617	-1.21851763	0.22610939	-1444.75912	345.95203	-1444.75912	345.95203
X Variable 1	65.3426144	110.315412	0.59232535	0.55507016	-153.72193	284.407159	-153.72193	284.407159
X Variable 2	2.51690026	0.9968464	2.52486267	0.01326771	0.53736088	4.49643965	0.53736088	4.49643965
X Variable 3	24.7773058	5.2556119	4.7144474	8.4884E-06	14.3407022	35.2139095	14.3407022	35.2139095
X Variable 4	0.10012296	0.31766683	0.31518231	0.75332936	-0.5307004	0.73094632	-0.5307004	0.73094632
X Variable 5	-1.19652171	6.52516591	-0.18337031	0.85490666	-14.154208	11.7611646	-14.154208	11.7611646
X Variable 6	22.6594869	8.58268673	2.64013911	0.00971915	5.61597203	39.7030018	5.61597203	39.7030018

Model 3:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.59534095
R Square	0.35443084
Adjusted R Sq	0.32724899
Standard Erro	59.3950984
Observations	100

ANOVA

	df	SS	MS	F	Significance F
Regression	4	183998.192	45999.5479	13.0392422	1.6733E-08
Residual	95	335138.883	3527.77771		
Total	99	519137.074			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-712.09486	457.07044	-1.5579543	0.12256935	-1619.4944	195.304685	-1619.4944	195.304685
log(Inc)	142.089558	109.047392	1.3030074	0.19572169	-74.396885	358.576002	-74.396885	358.576002
PriceLevel	1.93671083	0.99291058	1.95053901	0.05405903	-0.0344659	3.90788756	-0.0344659	3.90788756
GDP Growth	26.4010697	5.29589895	4.98519137	2.7801E-06	15.887381	36.9147584	15.887381	36.9147584
net flow	0.00025293	0.00030972	0.81665551	0.41616851	-0.0003619	0.0008678	-0.0003619	0.0008678

Model 4:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.63111738
R Square	0.39830915
Adjusted R Sq	0.3729748
Standard Erro	57.3410873
Observations	100

ANOVA

	df	SS	MS	F	Significance F
Regression	4	206777.047	51694.2617	15.7220977	6.5996E-10
Residual	95	312360.028	3288.00029		
Total	99	519137.074			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-516.44681	439.607951	-1.17478951	0.2430152	-1389.17894	356.285315	-1389.17894	356.285315
log(Inc)	58.8021842	108.066906	0.54412758	0.58762851	-155.73775	273.342118	-155.73775	273.342118
PriceLevel	2.43586444	0.96815756	2.51597937	0.01354743	0.51382868	4.3579002	0.51382868	4.3579002
GDP Growth	25.1966209	4.94697234	5.09334177	1.7805E-06	15.3756391	35.0176027	15.3756391	35.0176027
EconFreedom	23.109218	8.35874177	2.76467663	0.00684511	6.51501753	39.7034185	6.51501753	39.7034185

Model 5:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.48374782							
R Square	0.23401196							
Adjusted R Squ	0.20175983							
Standard Error	64.6979046							
Observations	100							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	121484.283	30371.0707	7.25570591	3.8361E-05			
Residual	95	397652.792	4185.81886					
Total	99	519137.074						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-170.945839	492.272318	-0.34725869	0.72916511	-1148.22995	806.338273	-1148.22995	806.338273
log(Inc)	-20.9259111	120.840586	-0.17316956	0.86288658	-260.824806	218.972984	-260.824806	218.972984
PriceLevel	3.42106847	1.08465342	3.154066	0.00215601	1.26775918	5.57437775	1.26775918	5.57437775
Population	0.24181156	7.01893092	0.03445134	0.97258953	-13.6925281	14.1761512	-13.6925281	14.1761512
EconFreedom	31.0259838	9.27013978	3.34687336	0.0011721	12.6224295	49.4295381	12.6224295	49.4295381

Model 6:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.62963							
R Square	0.39643393							
Adjusted R Sq	0.37757249							
Standard Erro	57.1304722							
Observations	100							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	205803.553	68601.1842	21.0182225	1.5029E-10			
Residual	96	313333.522	3263.89085					
Total	99	519137.074						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-281.96295	86.5704256	-3.2570355	0.00155673	-453.80389	-110.12202	-453.80389	-110.12202
PriceLevel	2.81959123	0.66088947	4.26635823	4.6589E-05	1.507736	4.13144646	1.507736	4.13144646
GDP Growth	24.8078511	4.87712485	5.08657291	1.805E-06	15.1268348	34.4888674	15.1268348	34.4888674
EconFreedom	24.2355527	8.06863307	3.00367515	0.00340159	8.2194426	40.2516627	8.2194426	40.2516627

TABLE G.3a 10% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Example: The 10% critical value for numerator $df = 2$ and denominator $df = 40$ is 2.44.

Source: This table was generated using the Stata® function invFtail.

TABLE G.3b 5% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator df (∞) is 2.37.

Source: This table was generated using the Stata® function invFtail.

TABLE G.3c 1% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator *df* = 3 and denominator *df* = 60 is 4.13.

Source: This table was generated using the Stata® function invFtail.

References

Bureau of Economic Analysis (2020). “GDP and Personal Income” Accessed March.15th, 2021 through https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=1&acrdn=8

Bureau of Economic Analysis (2020). “Personal Income by County, Metro, and Other Areas” Accessed March.15th, 2021 through https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas

Bureau of Economic Analysis (2020). “Regional Price Parities” Accessed March.15th, 2021 through https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=1&acrdn=8

Dennis R.Capozza, Patric H. Hendershott, Charlotte Mack, and Christopher J. Mayer(2002). “Determinants of Real House Price Dynamics”. Accessed through https://www.nber.org/system/files/working_papers/w9262/w9262.pdf

Federal Housing Finance Agency (2020). “100 Largest Metropolitan Statistical Areas (Seasonally Adjusted and Not Adjusted) ” Accessed March.15th, 2021 through https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx#qat

Gallin, Joshua (2006). "The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets," Real Estate Economics, vol. 34, no. 3, pp. 417-438.

Martijn I. Droes and Alex van de Minne(2017). “Do the Determinants of House Prices Change over Time? Evidence from 200 Years of Transactions Data”. Accessed Mar. 17th, 2021 through https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi7merBk8XvAhVLhq0KHQDLCkUQFjAGegQIBxAD&url=https%3A%2F%2Fwww.aeaweb.org%2Fconference%2F2017%2Fpreliminary%2Fpaper%2F7NHnQiSz&usg=AOvVaw1k3pL8TRgX1_Fu_-AxWWxa

United States Census Bureau (2020). “Metro-to-Metro Migration Flows” Accessed March.15th, 2021 through https://www.census.gov/topics/population/migration/guidance/metro-to-metro-migration-flows.html